

КАФТАНОВ А.Н.,

к.м.н., ООО «К-СКАЙ», Петрозаводск, Россия, e-mail: akaftanov@webiomed.ru

АНДРЕЙЧЕНКО А.Е.,

к.ф.-м.н., ООО «К-СКАЙ», Петрозаводск, Россия, e-mail: aandreychenko@webiomed.ru

ГУСЕВ А.В.,

к.т.н., ФГБУ «Центральный научно-исследовательский институт организации и информатизации здравоохранения» Минздрава России, Москва, Россия, e-mail: agusev@webiomed.ai

ОБЗОР МЕТОДИЧЕСКИХ ПОДХОДОВ К ОЦЕНКЕ КАЧЕСТВА ВЕДЕНИЯ ЭЛЕКТРОННЫХ МЕДИЦИНСКИХ КАРТ

DOI: 10.25881/18110193_2024_3_6

Аннотация. Переход на ведение электронных медицинских карт (ЭМК) является одним из базовых направлений цифровой трансформации здравоохранения. Одной из актуальных современных проблем ведения ЭМК является качество данных, которые накапливаются в современных медицинских информационных системах. Учитывая растущую роль ЭМК в качестве источника информации для систем поддержки принятия врачебных решений, внедрение элементов управления на основе первичных данных, а также развитие исследований в сфере данных реальной клинической практики (RWD), возрастает потребность в надежных и объективных методах оценки качества данных, накапливаемых в ЭМК. В этой связи разработка надежных методов и инструментов оценки качества данных (ОКД) в ЭМК является актуальной научной задачей.

Цель. Изучить и систематизировать предложенные в научной литературе подходы, методы и критерии ОКД ЭМК.

Материалы и методы. Были изучены обзоры и оригинальные работы по тематике ОКД ЭМК. Источники были выявлены в результате систематического поиска в четырех электронных библиографических базах данных: PubMed, Web of Science, Scopus и РИНЦ.

Результаты. В работе представлены основные подходы и критерии оценки качества данных ЭМК, проведена гармонизация терминов и определений ОКД, выделены ключевые компоненты, необходимые для внедрения системы ОКД ЭМК.

Заключение. Сформулированные в обзоре типовые критерии ОКД ЭМК могут быть использованы для дальнейших исследований и разработок инструментов ОКД, в том числе со стороны разработчиков медицинских информационных систем и организаторов здравоохранения, ответственных за цифровую трансформацию отрасли. Также данная работа поможет устранить путаницу в вопросах управления качеством данных ЭМК и предоставит руководство, необходимое для разработки эффективных программ для проведения ОКД.

Ключевые слова: электронная медицинская карта; ЭМК; качество данных; оценка качества данных.

Для цитирования: Кафтанов А.Н., Андрейченко А.Е., Гусев А.В. Обзор методических подходов к оценке качества ведения электронных медицинских карт. Врач и информационные технологии. 2024; 3: 6-19. doi: 10.25881/18110193_2024_3_6.

KAFTANOV A.N.,

PhD, K-Skai, Petrozavodsk, Russia, e-mail: akaftanov@webiomed.ru

ANDREYCHENKO A.E.,

PhD, K-Skai, Petrozavodsk, Russia, e-mail: aandreychenko@webiomed.ru

GUSEV A.V.,

PhD, Federal Research Institute for Health Organization and Informatics, Moscow, Russia,
e-mail: agusev@webiomed.ai

REVIEW OF METHODOLOGICAL APPROACHES TO ASSESSING THE QUALITY OF ELECTRONIC HEALTH RECORDS MANAGEMENT

DOI: 10.25881/18110193_2024_3_6

Abstract. Transition to electronic medical records (EMR) is one of the basic directions of digital transformation of healthcare. One of the urgent modern problems of EMR management is the quality of data that are accumulated in modern medical information systems. Given the growing role of EMRs as a source of information for medical decision support systems, the introduction of management elements based on primary data, and the development of research in the field of real-world clinical practice data (RWD), there is a growing need for reliable and objective methods to assess the quality of data accumulated in EMRs. In this regard, the development of reliable methods and tools for data quality assessment (DQA) in EMR is an urgent scientific task.

Aim. To study and systematize the approaches, methods and criteria of proposed in the scientific literature.

Materials and Methods. Reviews and original articles on the subject of EMRs DQA were studied. Sources were identified by systematic search in four electronic bibliographic databases: PubMed, Web of Science, Scopus and RSCI.

Results. The paper presents the main approaches and criteria for assessing the quality of EMR data, harmonizes the terms and definitions of DQA, and identifies the key components required to implement an EMRs DQA system.

Conclusion. The generic EMRs DQA criteria formulated in the review can be used for further research and development of DQA tools, including by medical information system developers and health care organizers responsible for the digital transformation of the industry. Also, this work will help eliminate confusion about EMR data quality management and provide the guidance needed to develop effective DQA programs.

Keywords: electronic medical record; EMR; data quality; data quality assessment.

For citation: Kaftanov A.N., Andreychenko A.E., Gusev A.V. Review of methodological approaches to assessing the quality of electronic health records management. Medical doctor and information technology. 2024; 3: 6-19. doi: 10.25881/18110193_2024_3_6.

ВВЕДЕНИЕ

В настоящее время медицинские информационные системы (МИС) широко внедрены в практическом здравоохранении России. Свыше 90% отечественных государственных и муниципальных медицинских организаций (МО) используют МИС различных производителей. Важнейшей задачей внедрения МИС является переход на ведение электронных медицинских карт (ЭМК), которые играют ключевую роль в цифровой трансформации здравоохранения, позволяя МО последовательно собирать, систематизировать и предоставлять доступ медицинским работникам к информации о лечебно-диагностических процессах, результатах обследования, лечения и наблюдения пациентов [1–3].

Повсеместное внедрение ЭМК привело к тому, что в российском здравоохранении накапливаются огромные объемы цифровых медицинских данных, которые являются оцифрованным артефактом реальной клинической практики, и обладают большим потенциалом для проведения широкого спектра клинико-практических исследований при относительно невысоких затратах [4, 5].

Накапливаемые в ЭМК данные становятся источником и драйвером для развития других направлений цифровой трансформации, таких как внедрение систем поддержки принятия врачебных решений (СППВР), созданных в том числе с использованием технологий искусственного интеллекта (ИИ), переход на управление на основе данных, исследования и принятие решений на основе данных реальной клинической практики (RWD) и т.д. [6].

Данные ЭМК представляют собой фактическую информацию о состоянии здоровья пациента, процессе его лечения и результатах лечения. Эти факты могут быть представлены в виде демографических данных пациента, диагнозов, протоколов врачебных осмотров, лабораторных, инструментальных и патологоанатомических исследований, выполненных процедур и операций, записей о лекарственных назначениях и рекомендациях и т.д. Как правило, медицинские работники регистрируют вышеперечисленную информацию в виде электронных медицинских записей с помощью форм и шаблонов. Степень структуризации записей зависит от реализации МИС и требований к ведению

медицинской документации в МО в регионе. Вместе с этим применяются и такие способы внесения данных в ЭМК, как автоматическая передача из медицинских изделий, сканирование бумажных носителей, а также средства преобразования речи в текст [7].

Известно, что с учетом многолетней истории исследований, практических разработок и, в итоге, внедрения ЭМК в практическое здравоохранение, качество данных (КД), собираемых в ЭМК, имеет различный уровень. В некоторых случаях, несмотря на формальное ведение ЭМК, накопленные в них сведения являются непригодными для повторного использования, включая работу СППВР или управления на основе данных.

Например, в ряде случаев отмечается отсутствие временных меток у записей, неподобные записи, значения, выходящие за пределы допустимых диапазонов, дубликаты и т.д. Проблемы с КД, особенно в ЭМК, являются следствием многих человеческих и технических факторов, таких как высокая загруженность медицинского персонала на работе, неточный сбор данных, ложные сведения от пациентов и небрежное ведение документации [8–10]. Другие факторы, такие как вариации клинической практики и отсутствие четких протоколов сбора данных, не интуитивный дизайн МИС, а также организационные факторы, такие как нарушение рабочего процесса, ротация персонала, разный уровень навыков работы с компьютером у персонала и неправильное использование, например, практика копирования и вставки, также непреднамеренно способствуют сбору некачественных медицинских данных [11–14].

Под КД подразумевается соответствие данных потребностям и ожиданиям потребителей, включая другие информационные системы, предназначенные для анализа и интерпретации данных из ЭМК. Таким образом, данные обладают высоким или низким качеством, если они, соответственно, пригодны или непригодны к последующему использованию. Следовательно, методология оценки КД ЭМК напрямую зависит от целей использования этих данных [15].

Отечественные исследователи давно обращают внимание на важность работ в области КД ЭМК [16–18]. Тем не менее, несмотря на некоторое количество публикаций, посвященных

проблемам в области КД и управления данными, направление ОКД в области здравоохранения в РФ развито недостаточно.

К сожалению, цена проблем с КД ЭМК как правило высока, что может негативно сказываться на безопасности пациентов, качестве лечения, эффективности управления ресурсами и достоверности клинических исследований. Сегодня многие медицинские ошибки имеют в качестве первопричины ошибки в данных [19]. Ошибки в данных также влияют на координацию медицинской помощи и ставят под угрозу операционную эффективность, затрудняя отслеживание успешности программ и реагирование на возникающие угрозы [20, 21].

Кроме того, клинические исследования и инструменты поддержки принятия решений, основанные на данных ЭМК, также несут большие затраты на подготовку данных, при этом все равно оставаясь в зоне высоких рисков получения недостоверных результатов [22–25]. Кроме того, растет объем непригодных для использования данных ЭМК. Учитывая критическую значимость этих проблем, связанных с КД, и высокую частоту повторного использования данных ЭМК, меры по обеспечению того, чтобы имеющиеся данные ЭМК были пригодны для дальнейшего анализа и соответствовали предполагаемым целям использования, являются крайне важными.

В предыдущих работах отмечалось, что обеспечение пригодности тех или иных данных для использования в целом включает четыре основных этапа: определение, измерение, анализ и совершенствование [26–28], как показано на рис. 1.

Первый этап – Определение: обычно фокусируется на определении контекста использования, элементов данных, представляющих интерес, проблем данных или измерений, которые необходимо исследовать.

Второй этап – Измерение: нужен для определения состояния КД набора данных. Обычно он включает в себя выявление проблем в наборе данных и составление отчета о состоянии набора данных на основе критериев. Результатом этапа измерения обычно является отбор записей с проблемами в данных и расчет метрик, отображающих степень выявленных проблем с данными в наборе.

Третий этап – Анализ: подразумевает оценку выявленных проблем, в том числе их влияния на результат использования данных.



Рисунок 1 — Типичная схема оценки и управления качеством данных.

Четвертый этап – Совершенствование: включает в себя мероприятия по улучшению или повышению пригодности данных для использования по назначению.

Во многих случаях об ошибках в данных редко сообщают или даже не признают их, когда они происходят. Между тем, оценка КД (ОКД), включающая четыре вышеперечисленных этапа, гарантирует что имеющиеся данные ЭМК являются полными, непротиворечивыми и пригодными для последующего использования. Кроме того, потребители данных ЭМК могут извлечь из ОКД множество потенциальных выгод, включая организацию мер по повышению качества ведения ЭМК, повышение эффективности инструментов сбора данных, снижение затрат на подготовку данных ЭМК к анализу, обеспечение четкой интерпретации результатов и углубление знаний о заболеваниях и методах лечения [19, 29].

В последнее десятилетие было предложено несколько методов ОКД ЭМК [30–33]. Однако многие организации не используют научно-обоснованные и универсальные методики ОКД ЭМК, предпочитая проводить эту работу в ручном режиме, в основном путем внутренних аудитов и аналитических обзоров, требующих значительных человеческих ресурсов и времени, что не подходит для больших объемов накопленных данных или ситуаций, когда время ограничено [34–36]. Кроме того, результаты таких специальных оценок трудно воспроизводимы и субъективны, поскольку они часто проводятся непоследовательно, а оценщики отличаются по уровню навыков и знаний [37, 38].

В связи с этим возникает потребность в надежных и автоматизированных инструментах ОКД ЭМК, основанных на комплексной и научно-обоснованной (основанной на научных публикациях в рецензируемых журналах) методике, позволяющей оценить влияние имеющихся в ЭМК проблем и управлять рисками использования данных ЭМК. Это требование становится еще более актуальным в настоящее время, когда все чаще звучат призывы к повышению прозрачности и доверия к управлению системой здравоохранения [19, 30, 39].

ЦЕЛИ ИССЛЕДОВАНИЯ

Целью данного обзора стало изучение и обобщение предложенных в научной литературе компонентов, необходимых для реализации процессов ОКД ЭМК.

МАТЕРИАЛЫ И МЕТОДЫ

Для поиска релевантных источников информации были использованы 4 электронные библиографические базы данных: PubMed, Web of Science, Scopus и РИНЦ. В поисковом запросе применялись строки «Data quality assessment of Electronic health records» и «Оценка качества данных электронных медицинских карт». Период поиска: все публикации до 01.03.2024.

Была обнаружено 37 публикаций. После анализа названий и резюме для пристального изучения отобрано 9 публикаций по тематике ОКД ЭМК.

РЕЗУЛЬТАТЫ

Одной из первых работ по теме исследования стала методика Weiskopf и др., которые в 2013 г. [30] предложили использовать 5 критериев качества ЭМК:

1. Полнота (completeness): наличие данных в ЭМК;
2. Корректность (correctness): правдивость данных в ЭМК;
3. Согласованность (concordance): соответствие между элементами внутри ЭМК и между ЭМК и другими источниками данных;
4. Правдоподобность (plausibility): степень, в которой данные ЭМК имеют смысл в широком медицинском контексте (не противоречат общепринятым медицинским знаниям);
5. Актуальность (currency): точность данных ЭМК в тот период времени, когда они были

записаны, и насколько актуальны сейчас эти данные [5, 30].

По данным обзора Lewis A. и др. (2023) именно критерии, предложенные Weiskopf и др., стали наиболее часто оцениваемыми критериями КД ЭМК в изученных авторами источниках: полнота (completeness), корректность (correctness), согласованность (concordance), актуальность (currency), правдоподобность (plausibility). Также авторы систематизировали способы оценки этих критериев, которые приведены далее. Полнота оценивалась путем сравнения с: 1) другим набором данных в рамках ЭМК или вне ее; 2) золотым стандартом. Для оценки корректности использовался метод согласования элементов данных. Согласованность оценивалась чаще всего с помощью взаимного анализа элементов данных или согласования источников данных. Актуальность чаще всего оценивалась с помощью просмотра журнала регистрации, но также использовались согласование элементов данных и сравнение распределений. Правдоподобность чаще всего оценивалась путем сравнения распределений. Авторы отметили также соответствие (conformance) и смещение (bias), или неслучайный пропуск, в качестве дополнительных критериев ОКД. Соответствие оценивалось с помощью структурного согласования (соответствие определенной структуре или формату данных). Смещение оценивалось чаще всего с помощью согласования элементов данных [40].

Kahn и др. (2016) [41] пришли к выводу, что текущая литература по ОКД ЭМК непоследовательна в использовании терминов, описывающих сложные многомерные аспекты КД. Отсутствие гармонизированных определений КД затрудняет понимание и сравнение результатов ОКД, полученных по различным методикам. Стандартизация терминов и методов, используемых для ОКД, а также показателей и форм отчетности, используемых для представления результатов ОКД, необходимы для улучшения понимания и повышения прозрачности собранных данных и результатов ОКД.

Авторы предприняли попытку стандартизации и гармонизации терминов и определений ОКД. Они разделили показатели измерения КД на три категории и среди них выделили подкатегории. Эти категории с примерами показаны в таблице 1.

Таблица 1 — Термины и определения ОКД с примерами [41]

| Верификация | | Валидация | |
|---|---|---|---|
| Определение | Пример | Определение | Пример |
| Соответствие: соответствуют ли значения данных указанным стандартам и форматам? | | | |
| Соответствие значений | | | |
| а. Значения данных соответствуют внутренним ограничениям формата. б. Значения данных соответствуют допустимым значениям или диапазонам. | а. Пол – это только один символ. б. Пол имеет только значения «М» или «Ж». | а. Значения данных соответствуют ограничениям, основанным на внешних стандартах. | а. Значения для основного языка соответствуют стандартам ISO. |
| Соответствие отношений | | | |
| а. Значения данных соответствуют ограничениям отношений. б. Уникальные (ключевые) значения данных не дублируются. с. Изменения в модели данных или версионирование модели данных. | а. Номер медицинской карты пациента связан с другими таблицами, если есть необходимость. б. Номер медицинской карты присваивается одному пациенту. с. Данные версии 1 не включают в себя время выписки. | а. Значения данных соответствуют ограничениям, основанным на внешних стандартах. | а. Значения данных соответствуют всем требованиям not null в общем формате обмена данными между различными учреждениями. |
| Соответствие вычислений | | | |
| а. Вычисленные значения соответствуют вычислительным или программным спецификациям. | а. Значения индекса массы тела (ИМТ) из базы данных и вычисленные с помощью прямого расчета, идентичны. | а. Вычисленные результаты, основанные на опубликованных алгоритмах, дают значения, совпадающие с проверочными значениями, предоставленными внешним источником. | а. Вычисленные перцентили ИМТ дают идентичные значения по сравнению с результатами тестов и предоставленными значениями. |
| Полнота: присутствуют ли значения данных? | | | |
| а. Отсутствие значений данных в конкретный момент времени соответствует ожиданиям. б. Отсутствие изменений значений данных со временем соответствует ожиданиям. | а. Переменная идентификатора имеет недостающие значения. б. Пол не должен быть равен null. с. Время медицинской выписки отсутствует в течение трех дней подряд. | а. Отсутствие значений данных в какой-то момент времени согласуется с проверенными эталонами или внешними источниками. б. Отсутствие изменения значений данных со временем согласуется с надежными эталонами или внешними источниками. | а. В текущей переменной идентификатора отсутствует в два раза больше значений, чем в валидированной базе данных. б. Снижение количества кодов МКБ-9 соответствует внедрению МКБ-10 |
| Правдоподобность: правдоподобны ли значения данных? | | | |
| Уникальность | | | |
| а. Значения данных, идентифицирующие один объект, не дублируются. | а. Пациенты из одного учреждения не имеют нескольких идентификаторов. | а. Значения данных, идентифицирующие один объект во внешнем источнике, не дублируются. | а. Идентификатор учреждения не относится к нескольким учреждениям. |

Таблица 1 — Термины и определения ОКД с примерами [41] (продолжение)

| Верификация | | Валидация | |
|--|--|---|--|
| Определение | Пример | Определение | Пример |
| Вневременное правдоподобие | | | |
| <p>а. Значения и распределения данных согласуются с внутренними измерениями или частными знаниями.</p> <p>б. Значения и распределения данных согласуются для независимых измерений.</p> <p>с. Логические отношения между значениями согласуются с частными или общепринятыми знаниями (включает «ожидаемые» пропуски).</p> <p>д. Значения повторных измерений одного и того же факта демонстрируют ожидаемую изменчивость.</p> | <p>а. Показатели роста и веса положительные.</p> <p>б. Количество уникальных пациентов по диагнозам соответствует ожиданиям</p> <p>с. Распределение приемов на одного пациента или распределение лекарств на один прием соответствует ожиданиям</p> <p>д. Измерение уровня глюкозы в сыворотке крови соответствует измерению уровня глюкозы с помощью тест-полоски.</p> <p>е. Оральная и аксиллярная температуры схожи.</p> <p>ф. Половые признаки согласуются с контекстом, специфическим для данного пола (беременность, рак простаты).</p> <p>г. Значения роста одинаковы при измерении двумя разными медсестрами в одном и том же учреждении с использованием одного и того же оборудования.</p> | <p>а. Значения и распределения данных (включая распределения по подгруппам) согласуются с проверенными эталонами или внешними источниками.</p> <p>б. Аналогичные значения для идентичных измерений получены из двух независимых баз данных, представляющих одни и те же наблюдения с одинаковой достоверностью.</p> <p>с. Две зависимые базы данных (например, база данных 1 и база данных 2, дают одинаковые значения для идентичных переменных.</p> | <p>а. Показатели HbA1c, полученные в больнице и референс-лаборатории, статистически схожи при одинаковых условиях.</p> <p>б. Распределение пациентов с диагнозами сердечно-сосудистых заболеваний соответствует показателям CDC* для тех же возрастных и половых групп</p> <p>с. Коды диабета по МКБ-9 совпадают в двух независимых базах данных, обслуживающих схожие группы населения.</p> <p>д. Зафиксированная дата рождения соответствует регистрационным данным одного и того же пациента.</p> |
| Временное правдоподобие | | | |
| <p>а. Наблюдаемые или полученные значения соответствуют ожидаемым временным характеристикам.</p> <p>б. Последовательности значений, представляющие переходы состояний, соответствуют ожиданиям.</p> <p>с. Показатели плотности значений данных по отношению к знаменателю, ориентированному на время, будут основаны на внутренних знаниях.</p> | <p>а. Дата поступления меньше даты выписки.</p> <p>б. Дата первичной вакцинации предшествует дате ревакцинации.</p> <p>с. Аналогичное количество наблюдений за пациентами между циклами извлечение-преобразование-загрузка.</p> <p>д. Количество посещений отделений неотложной помощи по месяцам показывает ожидаемый всплеск в сезон гриппа.</p> <p>е. Количество приемов лекарств на пациента в день соответствует ожидаемому.</p> | <p>а. Наблюдаемые или полученные значения имеют схожие временные характеристики по одному или нескольким внешним золотым стандартам.</p> <p>б. Последовательность значений, которые представляют собой переходы состояний, похожи на внешние золотые стандарты.</p> <p>с. Ожидается, что показатели плотности значений данных по отношению к знаменателю, ориентированному на время, будут основаны на внешних знаниях.</p> | <p>а. Продолжительность пребывания по типам амбулаторных процедур соответствует данным Medicare для аналогичных групп населения.</p> <p>б. Последовательность вакцинаций соответствует рекомендациям CDC.</p> <p>с. Подсчет количества посещений отделения неотложной помощи по месяцам показывает всплеск в сезон гриппа и совпадает с данными местных отделов здравоохранения.</p> <p>д. Лекарства на один пациент-день соответствуют заявленным данным.</p> |

Примечание: * – Centers for Disease Control and Prevention.

Категория «Соответствие» (conformance) фокусируется на критериях ОКД, которые описывают соответствие представления данных внутреннему или внешнему формату, синтаксическим или структурным ограничениям. Данная категория соответствия часто используется в документе, называемом «словарем данных», в котором описан предполагаемый формат и допустимые значения для каждого элемента данных. Соответствие разделено на три подкатегории: соответствие значений, соответствие отношений и соответствие вычислений. Соответствие значений определяет, соответствуют ли записанные элементы данных заранее определенной архитектуре, управляемой ограничениями. Соответствие отношений определяет, соответствуют ли записанные элементы данных дополнительным структурным ограничениям, налагаемым физической структурой базы данных, в которых хранятся значения данных, например, таким как ограничение на возможность полям данных принимать значение null, отношениям первичного и внешнего ключа. Соответствие вычислений определяет, используются ли вычисления для создания производных переменных из существующих, и соответствуют ли выходные значения вычислений техническим и функциональным спецификациям.

Категория «Полнота» (completeness) фокусируется на характеристиках, которые описывают частоту появления атрибутов данных в наборе данных без привязки к их значениям. Показатели полноты оценивают отсутствие данных в одном или нескольких моментах во времени, без привязки к их структуре или правдоподобию, которые оцениваются в категориях соответствия и правдоподобия.

Категория «Правдоподобие» (plausibility) фокусируется на характеристиках, которые описывают правдоподобность значений данных. Для этой категории правдоподобие определяется значением переменной, помещенной в контекст других переменных или во временную последовательность (т.е., например, лечению пациента по поводу заболевания должен предшествовать соответствующий диагноз). В отличие от соответствия и полноты, которые фокусируются только на структуре и наличии переменных, правдоподобие фокусируется на фактических значениях переменных, их

распределении и взаимосвязи с другими переменными. Правдоподобие подразумевает наличие приемлемого диапазона значений переменных и их распределения, а не требует подтверждения точного значения. Правдоподобие имеет подкатегории уникальности, вневременности и временности. Подкатегория уникальности предназначена для определения того, являются ли объекты (сущности, наблюдения, факты) в данных несколько раз. Дублирование часто возникает, когда объединяются разрозненные потоки данных или вследствие ошибок извлечения данных. Вневременная подкатегория определяет, согласуются ли наблюдаемые значения данных и их распределения с «общепринятыми» знаниями, надежными внешними источниками или золотым стандартом. Примеры включают физические величины, которые не могут быть отрицательными, не могут превышать возможные границы или (в медицине) представлять физиологически невозможные состояния. Вневременная подкатегория также фокусируется на наблюдаемых взаимосвязях между независимыми переменными, которые должны иметь ожидаемые или известные взаимосвязи. Временная подкатегория направлена на определение того, как изменяются значения переменных во времени, и происходит ли это ожидаемым образом (сохраняются ли значения с течением времени, как ожидалось, происходят ли последовательности событий, как ожидалось). В медицинских данных наблюдаемыми временными свойствами могут быть временные циклы, такие как суточные колебания, циклические схемы лечения, рецидивы, повторяющиеся вспышки заболеваний [41].

В отличие от Weiskopf и Lewis, Kahn дает более развернутые определения критериев оценки КД. Категории «Соответствие» и «Полнота» у Kahn и Lewis аналогичны. Категория «Правдоподобие» у Kahn соответствует категориям «Корректность», «Правдоподобие» и «Актуальность» у Lewis. У Kahn нет категорий «Смещение» и «Согласованность». Категория «Смещение» у Lewis ближе всего к категории «Полнота» у Kahn. В то же время категорию «Согласованность» у Lewis трудно сопоставить с какой-либо категорией у Kahn.

Ozonze O. и др. (2023) определили характеристики, необходимые для внедрения системы ОКД ЭМК, которые были сгруппированы в пять

категорий [42]. Первые четыре категории соответствуют схеме на рисунке 1:

1. Определение задач ОКД (КД-задача);
2. Получение измерений, используемых для оценки определенных задач ОКД (КД-измерения);
3. Сбор и обработка целевых данных (КД-данные);
4. Механизмы реализации мер (КД-механизмы);
5. Механизмы фиксации результатов (КД-отчет).

КД-задача описывает требования для проведения ОКД, которые могут быть определены заинтересованными сторонами, внутренними или внешними по отношению к человеку или организации, осуществляющей ОКД, такими как потребители данных, разработчики программ, производители данных и др. Типичные элементы КД-задачи включают требования, непосредственно связанные с качеством, такие как измерения для исследования (КД- измерения), элементы данных, представляющие интерес, и метрики для определения того, достаточно ли хорош набор данных для предполагаемого случая использования (КД-метрики). Требования КД-задачи могут также включать, например, как она должна выполняться, периодичность, с которой она должна выполняться, должна ли она проводиться по требованию или в определенные промежутки времени, например, в конце каждого дня. Как правило, КД-задача направлена на оценку одного или нескольких КД-измерений в заданном наборе данных. Как было отмечено в данном исследовании [42], определения этих КД-измерений часто варьируются и берутся из различных источников, включая экспертов в данной области, обзоры литературы и т.д.

Выбор КД-измерений для оценки – еще один критический момент при постановке КД-задач, поскольку определенные КД-измерения могут быть применимы только для элементов данных определенной области и типа данных. Аналогичным образом маловероятно, что программы ОКД смогут оценить все возможные КД-измерения по всем доступным элементам данных особенно для больших наборов данных ЭМК. Некоторые требуемые КД-измерения могут быть недоступны или слишком сложны для внедрения. Кроме того, наборы данных с большим количеством атрибутов, сложными типами

данных, такими как изображения, и большими размерами могут потребовать больше ресурсов, чем позволяет доступный КД-механизм. Поэтому в таких сценариях для повышения эффективности ОКД необходимо найти компромисс между КД-измерениями, элементами данных, временем и возможностями КД-механизма. Примерами таких компромиссов могут быть фокусировка на элементах данных, необходимых для предполагаемых сценариев использования или имеющих больший вес в отношении их вклада в общее качество набора данных.

В зависимости от поставленной цели КД-задача может включать метрику и порог для определения того, достаточно ли хорош целевой набор данных для предполагаемого использования. Эта метрика (КД-метрика), представляющая собой совокупный балл, может быть количественной (например, количество), простое соотношение (например, процент), категориальной (например, ординальной, логической) или другой сложной метрикой [42].

КД-измерение обычно включает одно или несколько сравнений, касающихся содержания элементов данных, их происхождения, свойств (например, типа, формата) и референсных элементов на разных уровнях данных (ячейка, запись, таблица). Референсными элементами могут быть любые значения, содержащиеся в других элементах данных в том же наборе, результаты других КД-измерений и явно определенные значения, такие как числа, логические выражения, текст, диапазоны значений и регулярные выражения. Тип данных оцениваемого элемента данных может определять тип требуемого референсного элемента. Например, для оценки элементов данных типа «число», скорее всего, будет использоваться проверка диапазона.

Обычные сравнения включают оценку соответствия значений, например, наличие значений, соответствие определенным характеристикам, точность, допустимые диапазоны, функциональные зависимости и причинно-следственные связи. Они также могут включать оценку согласованности с другими источниками данных, такими как предыдущая копия тех же данных, а также другие наборы данных в одной или разных МИС. При проведении КД-измерений с использованием разрозненных наборов данных необходимо

учитывать, что наборы данных могут иметь синтаксические и семантические различия. И хотя существуют различные функции преобразования и инструменты для нормализации наборов данных, чрезмерные преобразования могут завысить или занижить оценку КД-измерений.

Кроме того, как уже говорилось ранее, КД-измерения применяются к определенным уровням данных (ячейка, поле, запись, таблица). Например, при оценке соответствия значений КД-измерения могут быть направлены на отдельные ячейки данных в записях, в частности проверка соответствия отдельных ячеек данных таким требованиям, как тип и формат данных. Аналогичным образом, некоторые КД-измерения применяются на уровне поля, сравнивая результаты агрегирования выбранных наблюдений (записей) в этом поле с эталонной информацией, например, выявление одновариантных выбросов и оценка избыточности. Другие методы предполагают использование нескольких элементов данных на уровне записей, например, выявление нарушений функциональной зависимости и согласование нескольких переменных (как вариант, полей, содержащих диагнозы и концепции лечения). Кроме того, несколько элементов данных можно сравнивать между агрегированными записями, например, сравнивая значение элемента данных с последовательными значениями тех же полей для данного субъекта, чтобы определить, не изменились ли значения с течением времени. Также возможно, чтобы КД-измерения применялись на уровне таблицы, и чтобы несколько КД-измерений объединялись с помощью логических связей типа AND, OR и NOT для проведения сложных КД-измерений [42].

Еще одним фактором является размер набора данных, поскольку он не всегда позволяет

своевременно, экономично или безопасно проверять каждую запись в наборе. Поэтому вместо оценки всего набора данных можно выбрать подмножество исходных данных, используя стратегии выборки и рандомизации [29, 43]. Однако результаты оценки не всегда отражают КД-статус набора данных. Кроме того, определение подходящего размера выборки данных, достаточной для оценки состояния всего набора данных, может оказаться сложной задачей [43].

Данные ЭМК не всегда имеют одинаковую структуру и в таких случаях может потребоваться трансформация данных для преобразования исходных данных, особенно неструктурированных, в формат, подходящий для целевых КД-измерений [26, 44].

КД-отчет обеспечивает обратную связь, позволяющую заинтересованным сторонам оценивать свои наборы данных, и может включать рекомендации по улучшению КД. Типичный КД-отчет может содержать набор записей, удовлетворяющих оцениваемым КД-измерениям, баллы КД-метрик и другие данные, например, рекомендации по улучшению КД. Эти результаты могут быть доведены до сведения заинтересованных сторон с использованием предпочтительных методов предоставления информации, таких как таблицы и графики, которые позволяют быстро провести анализ. КД-отчеты также могут быть экспортированы или сохранены для последующего анализа.

КД-механизм – это программа, процесс или инструмент, используемый для реализации различных видов деятельности, связанных с выполнением КД-задачи. Общие характеристики КД-механизма были сгруппированы по категориям возможности настройки, удобства использования, масштабируемости, производительности и безопасности, как показано ниже в таблице 2 [42].

Таблица 2 — Пример характеристик КД-механизма [42]

| Характеристика | Описание |
|------------------------|---|
| Возможность настройки | Позволяет пользователям персонализировать, адаптировать или расширить процесс ОКД в соответствии с их требованиями или средой |
| Удобство использования | Позволяет пользователям эффективно и результативно выполнять задачи |
| Масштабируемость | Позволяет системе устойчиво реагировать на изменения в количестве доступных ресурсов, наборах данных или окружении |
| Производительность | Обеспечение удовлетворительного уровня отклика и стабильности системы при заданных рабочих нагрузках |
| Безопасность | Позволяет удовлетворить требования по безопасности, такие как конфиденциальность и защита прав собственности |

В отличие от Weiskopf, Lewis и Kahn, в части критериев КД (КД-измерений) Ozonze [42] делает акцент не на терминологии и определениях, а на выборе критерия в зависимости от задачи, типа и уровня данных.

ОБСУЖДЕНИЕ

Анализ литературы показал, что организации, использующие данные ЭМК, все больше осознают присущие им проблемы качества. Он также подтверждает растущее внимание к автоматизации ОКД ЭМК, вызванное главным образом необходимостью помочь исследователям подготовить данные ЭМК для решения исследовательских задач. Однако в настоящее время лишь несколько программ ОКД направлены на повышение КД в исходных ЭМК, что крайне важно для предотвращения медицинских ошибок и улучшения электронной документации. Последнее можно объяснить тем, что имеющиеся программы ОКД не настолько надежны, как хотелось бы, и сосредоточены на таких аспектах КД, как полнота и соответствие значений, которые можно считать тривиальными для реализации и которые в настоящее время поддерживаются различными инструментами интеграции и анализа данных. Кроме того, неструктурированные форматы данных, такие как текст и изображения, которые составляют большую часть данных, хранящихся в ЭМК [45], являются более сложными для анализа с точки зрения вычислений [46] и поэтому редко поддерживаются. Аналогичным образом, многие из предложенных методик ОКД тесно связаны с существующей инфраструктурой и доступны только для пользователей одной организации или какой-то одной системы ведения ЭМК. Некоторые из них также слишком техничны, лишены интерактивности и для работы с инструментом ОКД требуют от пользователей знаний о хост-системах и дополнительных языках программирования, например, знания R [47, 48]. Кроме того, они не проходят адекватной валидации, а значит не готовы для общего клинического использования.

Эти ограничения еще больше подчеркивают сложность разработки методики оценки качества ЭМК. Не было выявлено недостатка в работах, в которых обсуждаются теории ОКД, лучшие практики и другие концепции, связанные с

ОКД. Например, в нескольких концепциях, таких как концепция Total Data Quality Management (TDQM), описаны лучшие практики для улучшения общего КД и проведения ОКД с общей точки зрения [26, 27, 49–52], а в некоторых других – непосредственно для данных ЭМК [30, 39, 53, 54]. Однако остается неясным, как многочисленные теоретические концепции могут быть реализованы на практике. Например, многие существующие работы сосредоточены на стандартизации параметров КД и определении потенциальных методов оценки, но они не дают четкого представления о том, как эти методы могут быть внедрены в реальных условиях. Кроме того, лишь немногие исследования посвящены другим важнейшим аспектам ОКД: управлению данными [33, 43, 55], отчетности и применению результатов [56, 57]. Концепции также обсуждаются изолированно и, следовательно, содержат конкурирующие и неоднозначные термины, что вносит путаницу и затрудняет их практическое применение [58, 59].

ЗАКЛЮЧЕНИЕ

Данные ЭМК являются важнейшим компонентом современной системы здравоохранения и должны быть достаточно качественными, чтобы поддерживать лечебный процесс, а также быть пригодными для повторного использования, включая работу СППВР, применение технологий ИИ и управления на основе данных. Для обеспечения этого были предложены различные стратегии, в том числе ОКД ЭМК, для выявления и устранения проблем в данных. Тем не менее, в настоящий момент отсутствуют комплексные методики и инструменты для получения надежных и воспроизводимых оценок КД ЭМК, как минимум для РФ. В свете этого мы изучили литературу, чтобы устранить этот пробел и определить важные аспекты для разработки и внедрения новых методов и инструментов ОКД ЭМК.

Полученные нами результаты свидетельствуют о том, что автоматизация оценки качества ЭМК должна быть построена на основе адекватной и пригодной для сравнений методики, которая в свою очередь может быть разработана с учетом сделанных ранее в литературе предложений. Однако, как представляется, в целом отсутствует ясность в отношении процессов

ОКД ЭМК, что обусловлено контекстуальным характером требований к ОКД, неоднородностью данных ЭМК и сложностью разработки измерений для проверки наборов данных. Еще более тревожным является то, что качество процесса ОКД ЭМК неизвестно, и нет обязательств по сертификации того, что инструменты ОКД измеряют то, для чего они предназначены. Кроме того, растет спрос на универсальные проверки, применимые в различных контекстах. Организациям в сфере здравоохранения, которые планируют разработку и применение инструментов

ОКД ЭМК, будет полезен этот обзор, поскольку мы обобщили существующие компоненты, необходимые для реализации процессов ОКД. Мы также ожидаем, что эта работа поможет устранить путаницу в вопросах управления КД ЭМК и предоставит руководство, необходимое для разработки эффективных программ для проведения ОКД.

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов (The authors declare no conflict of interest).

ЛИТЕРАТУРА/REFERENCES

1. Warren LR, Clarke J, Arora S, et al. Improving data sharing between acute hospitals in England: an overview of health record system distribution and retrospective observational analysis of inter-hospital transitions of care. *BMJ Open* 2019; 9: e031637. doi: 10.1136/bmjopen-2019-031637.
2. Atasoy H, Greenwood BN, McCullough JS. The Digitization of Patient Care: A Review of the Effects of Electronic Health Records on Health Care Quality and Utilization. *Annu Rev Public Health*. 2019; 40: 487-500. doi: 10.1146/annurev-publhealth-040218-044206.
3. Häyrynen K, Saranto K, Nykänen P. Definition, structure, content, use and impacts of electronic health records: a review of the research literature. *Int J Med Inform*. 2008; 77(5): 291-304. doi: 10.1016/j.ijmedinf.2007.09.001.
4. Meystre SM, Lovis C, Bürkle T et al. Clinical Data Reuse or Secondary Use: Current Status and Potential Future Progress. *Yearb Med Inform*. 2017; 26(1): 38-52. doi: 10.15265/IY-2017-007.
5. Goldstein BA, Navar AM, Pencina MJ, Ioannidis JP. Opportunities and challenges in developing risk prediction models with electronic health records data: a systematic review. *J Am Med Inform Assoc*. 2017; 24(1): 198-208. doi: 10.1093/jamia/ocw042.
6. Topol E. The Topol Review Preparing the Healthcare Workforce to Deliver the Digital Future. 2019: 1-48.
7. Vuokko R, Mäkelä-Bengs P, Hyppönen H, Doupi P. Secondary use of structured patient data: interim results of a systematic review. *Stud Health Technol Inform*. 2015; 210: 291-5.
8. Collins SA, Bakken S, Vawdrey DK, et al. Clinician preferences for verbal communication compared to EHR documentation in the ICU. *Appl Clin Inform*. 2011; 2(2): 190-201. doi: 10.4338/ACI-2011-02-RA-0011.
9. Salomon RM, Blackford JU, Rosenbloom ST et al. Openness of patients' reporting with use of electronic records: psychiatric clinicians' views. *J Am Med Inform Assoc*. 2010; 17(1): 54-60. doi: 10.1197/jamia.M3341.
10. Peivandi S, Ahmadian L, Farokhzadian J, Jahani Y. Evaluation and comparison of errors on nursing notes created by online and offline speech recognition technology and handwritten: an interventional study. *BMC Med Inform Decis Mak*. 2022; 22(1): 96. doi: 10.1186/s12911-022-01835-4.
11. Colin NV, Cholan RA, Sachdeva B et al. Understanding the Impact of Variations in Measurement Period Reporting for Electronic Clinical Quality Measures. *EGEMS (Wash DC)*. 2018; 6(1): 17. doi: 10.5334/egems.235.
12. Bowman S. Impact of electronic health record systems on information integrity: quality and safety implications. *Perspect Health Inf Manag*. 2013; 10(Fall): 1c.
13. O'Donnell HC, Kaushal R, Barrón Y, et al. Physicians' attitudes towards copy and pasting in electronic note writing. *J Gen Intern Med*. 2009; 24(1): 63-8. doi: 10.1007/s11606-008-0843-2.
14. Coleman N, Halas G, Peeler W, et al. From patient care to research: a validation study examining the factors contributing to data quality in a primary care electronic medical record database. *BMC Fam Pract*. 2015; 16: 11. doi: 10.1186/s12875-015-0223-z.
15. ДАМА-DMBOK: Свод знаний по управлению данными. 2-е изд. 2020. Dama International [пер. с англ. Г. Агафонова]. М.: Олимп-Бизнес, 2020. 828 с.: ил. [DAMA-DMBOK: Svod znanij po upravleniyu dannymi. 2-e izd. 2020. Dama International [per. s angl. G. Agafonova]. М.: Олимп-Бизнес, 2020. 828 p.: il. (In Russ.)]

16. Любичын В.Н. Повышение качества данных в контексте современных аналитических технологий // Вестник ЮУрГУ. Серия: Компьютерные технологии, управление, радиоэлектроника. – 2012. – №23. [Ljubicyн VN. Povyshenie kachestva dannyh v kontekste sovremennyh analiticheskikh tehnologij. Vestnik JuUrGU. Serija: Komp'juternye tehnologii, upravlenie, radioelektronika. 2012; 23. (In Russ.)]
17. Килимова А.Д. Потоки данных в легкой промышленности // Компетентность. – 2022. – №3. [Kilimova AD. Potoki dannyh v legkoj promyshlennosti. Kompetentnost'. 2022; 3. (In Russ.)]
18. Афанасьев А.А., Кудинов В.А. Использование онтологического подхода для извлечения ожиданий к качеству данных корпоративных хранилищ // Экономика. Информатика. – 2022. – №49(3). – С.566-574. [Afanas'ev AA, Kudinov VA. Ispol'zovanie ontologicheskogo podhoda dlja izvlechenija ozhidaniy k kachestvu dannyh korporativnyh hranilishh. Jekonomika. Informatika. 2022; 49(3): 566-574. (In Russ.)] doi: 10.52575/2687-0932-2022-49-3-566-574.
19. Elliott RA, Camacho E, Jankovic D, et al. Economic analysis of the prevalence and clinical and economic burden of medication error in England. *BMJ Qual Saf.* 2021; 30(2): 96-105. doi: 10.1136/bmjqs-2019-010206.
20. Zozus MN, Penning M, Hammond WE. Factors impacting physician use of information charted by others. *JAMIA Open.* 2019; 2(1): 107-114. doi: 10.1093/jamiaopen/ooy041.
21. Munyisia EN, Reid D, Yu P. Accuracy of outpatient service data for activity-based funding in New South Wales, Australia. *Health Inf Manag.* 2017; 46(2): 78-86. doi: 10.1177/1833358316678957.
22. Kaplan B. How Should Health Data Be Used? *Camb Q Healthc Ethics.* 2016; 25(2): 312-29. doi: 10.1017/S0963180115000614.
23. Nouraei SA, Virk JS, Hudovsky A, et al. Accuracy of clinician-clinical coder information handover following acute medical admissions: implication for using administrative datasets in clinical outcomes management. *J Public Health (Oxf).* 2016; 38(2): 352-62. doi: 10.1093/pubmed/fdv041.
24. Feldman K, Faust L, Wuet X, et al. Beyond volume: The impact of complex healthcare data on the machine learning pipeline *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics).* 2017; 10344 LNAI: 150-169.
25. Hanauer DA, Mei Q, Vydiswaran VGV, et al. Complexities, variations, and errors of numbering within clinical notes: the potential impact on information extraction and cohort-identification. *BMC Med Inform Decis Mak.* 2019; 19(3): 75. doi: 10.1186/s12911-019-0784-1.
26. Batini C, Francalanci C, Cappiello C, Maurino A. Methodologies for data quality assessment and improvement. *ACM computing surveys (CSUR).* 2009; 41(3): 16.
27. Wang RY. A product perspective on total data quality management. *Communications of the ACM.* 1998; 41(2): 58-66. doi: 10.1145/269012.269022.
28. Veiga AK, Saraiva AM, Chapman AD, et al. A conceptual framework for quality assessment and management of biodiversity data. *PLoS One.* 2017; 12(6): e0178731. doi: 10.1371/journal.pone.0178731.
29. WHO, Data Quality Assessment of National and Partner Hiv Treatment and Patient Monitoring Systems. 2018. August: 1-68.
30. Weiskopf NG, Weng C. Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research. *J Am Med Inform Assoc.* 2013; 20(1): 144-51. doi: 10.1136/amiajnl-2011-000681.
31. Feder SL. Data Quality in Electronic Health Records Research: Quality Domains and Assessment Methods. *West J Nurs Res.* 2018; 40(5): 753-766. doi: 10.1177/0193945916689084.
32. Reimer AP, Milinovich A, Madigan EA. Data quality assessment framework to assess electronic medical record data for use in research. *Int J Med Inform.* 2016; 90: 40-7. doi: 10.1016/j.ijmedinf.2016.03.006.
33. Kahn MG, Raebel MA, Glanz JM, et al. A pragmatic framework for single-site and multisite data quality assessment in electronic health record-based clinical research. *Med Care.* 2012; 50(0): S21-9. doi: 10.1097/MLR.0b013e318257dd67.
34. Muthee V, Bochner AF, Osterman A, et al. The impact of routine data quality assessments on electronic medical record data quality in Kenya. *PLoS One.* 2018; 13(4): e0195362. doi: 10.1371/journal.pone.0195362.
35. Yadav S, Kazanji N, K C N, Paudel S, et al. Comparison of accuracy of physical examination findings in initial progress notes between paper charts and a newly implemented electronic health record. *J Am Med Inform Assoc.* 2017; 24(1): 140-144. doi: 10.1093/jamia/ocw067.

36. Abiy R, Gashu K, Asemaw T, et al. A Comparison of Electronic Medical Record Data to Paper Records in Antiretroviral Therapy Clinic in Ethiopia: What is affecting the Quality of the Data? *Online J Public Health Inform.* 2018; 10(2): e212. doi: 10.5210/ojphi.v10i2.8309.
37. Maletic JI, Marcus A, Data Cleansing: Beyond Integrity Analysis Iq, 2000: 1-10.
38. Daymont C, Ross ME, Russell Localio A, et al. Automated identification of implausible values in growth data from pediatric electronic health records. *J Am Med Inform Assoc.* 2017; 24(6): 1080-1087. doi: 10.1093/jamia/ocx037.
39. Brown JS, Kahn M, Toh S. Data quality assessment for comparative effectiveness research in distributed data networks. *Med Care.* 2013; 51(8S3): S22-9. doi: 10.1097/MLR.0b013e31829b1e2c.
40. Lewis AE, Weiskopf N, Abrams ZB, et al. Electronic health record data quality assessment and tools: a systematic review. *J Am Med Inform Assoc.* 2023; 30(10): 1730-1740. doi: 10.1093/jamia/ocad120.
41. Kahn MG, Callahan TJ, Barnard J, et al. A Harmonized Data Quality Assessment Terminology and Framework for the Secondary Use of Electronic Health Record Data. *EGEMS (Wash DC).* 2016; 4(1): 1244. doi: 10.13063/2327-9214.1244.
42. Ozonze O, Scott PJ, Hopgood AA. Automating Electronic Health Record Data Quality Assessment. *J Med Syst.* 2023; 47(1): 23. doi: 10.1007/s10916-022-01892-2.
43. Pageler NM, Grazier G'Sell MJ, Chandler W, et al. A rational approach to legacy data validation when transitioning between electronic health record systems. *J Am Med Inform Assoc.* 2016; 23(5): 991-4. doi: 10.1093/jamia/ocv173.
44. Ferrão JC, Oliveira MD, Janela F, Martins HM. Preprocessing structured clinical data for predictive modeling and decision support. A roadmap to tackle the challenges. *Appl Clin Inform.* 2016; 7(4): 1135-1153. doi: 10.4338/ACI-2016-03-SOA-0035.
45. Safran C. Update on Data Reuse in Health Care. *Yearb Med Inform.* 2017; 26(1): 24-27. doi: 10.15265/IY-2017-013.
46. Jensen PB, Jensen LJ, Brunak S. Mining electronic health records: towards better research applications and clinical care. *Nat Rev Genet.* 2012; 13(6): 395-405. doi: 10.1038/nrg3208.
47. Estiri H, Klann JG, Weiler SR, et al. A federated EHR network data completeness tracking system. *J Am Med Inform Assoc.* 2019; 26(7): 637-645. doi: 10.1093/jamia/ocz014.
48. User V, DeFalco FJ, Schuemie M, et al. Multisite Evaluation of a Data Quality Tool for Patient-Level Clinical Data Sets. *EGEMS (Wash DC).* 2016; 4(1): 1239. doi: 10.13063/2327-9214.1239.
49. Pipino LL, Lee YW, Wang RY. Data Quality Assessment Communications of the ACM. 2002; 45(4): 211. doi: 10.1145/505248.506010.
50. Naumann F, Rolker C. Assessment Methods for Information Quality Criteria Information Systems. 2000: 148-162. doi: 10.18452/9207.
51. Woodall P, Oberhofer M, Borek A. A classification of data quality assessment and improvement methods. *International Journal of Information Quality,* 2014. 3(4): 298-321. doi: 10.1504/IJIQ.2014.068656.
52. DAMA UK Working Group. The six primary dimensions for data quality assessment: defining data quality dimensions. 2013.
53. Weiskopf NG, Bakken S, Hripcsak G, Weng C. A Data Quality Assessment Guideline for Electronic Health Record Data Reuse. *EGEMS (Wash DC).* 2017; 5(1): 14. doi: 10.5334/egems.218.
54. Johnson SG, Speedie S, Simon G, et al. A Data Quality Ontology for the Secondary Use of EHR Data. *AMIA Annu Symp Proc.* 2015; 2015: 1937-46.
55. Hartzema AG, Reich CG, Ryan PB, et al. Managing data quality for a drug safety surveillance system. *Drug Saf.* 2013; 36(1): S49-58. doi: 10.1007/s40264-013-0098-7.
56. Kahn MG, Brown JS, Chun AT, et al. Transparent reporting of data quality in distributed data networks. *EGEMS (Wash DC).* 2015; 3(1): 1052. doi: 10.13063/2327-9214.1052.
57. Callahan T, Barnard J, Helmkamp L, et al. Reporting Data Quality Assessment Results: Identifying Individual and Organizational Barriers and Solutions. *EGEMS (Wash DC).* 2017; 5(1): 16. doi: 10.5334/egems.214.
58. Roomaney RA, Pillay-van Wyk V, Awotiwon OF, et al. Availability and quality of routine morbidity data: review of studies in South Africa. *J Am Med Inform Assoc.* 2017; 24(e1): e194-e206. doi: 10.1093/jamia/ocw075.
59. Estiri H, Stephens KA, Klann JG, Murphy SN. Exploring completeness in clinical data research networks with DQe-c. *J Am Med Inform Assoc.* 2018; 25(1): 17-24. doi: 10.1093/jamia/ocx109.